# Averaging techniques for single-trial analysis of oddball event-related potentials

Nanying Liang, Laurent Bougrain

# Averaging techniques for single-trial analysis of oddball event-related potentials

Nanying Liang, Laurent Bougrain

INRIA Nancy - Grand Est, France

www.loria.fr/~liangnna

**Abstract**

More and more effort is done in BCI research to improve its usability for patients, with respect to its communication speed and transmission accuracy. In this contribution, we experiment with BCI speller based on P300 evoked potential. More precisely, the typical form of event-related potential (ERP) inspires us to devise classification methods based on the similarity/dissimilarity in the time domain between single trials and one or several estimated ERP templates derived from subject recordings. The reliable estimation of template is difficult in a single trial due to the low signal-to-noise ratio (SNR) of electroencephalographic (EEG) signals. We first explicitly estimate the template using several averaging techniques: point-to-point averaging, cross-correlation alignment and dynamic time warping. Then we inexplicitly estimate several ERP templates using learning vector quantization algorithm combined with an extreme learning machine. Finally classification is realized based on the similarity/dissimilarity between the single trials and the template. Simulation is carried out using a BCI competition III data set acquired with the P300 speller paradigm. The experiments show that template-based classifiers can also obtain high accuracy.

## 1 Introduction

Brain-computer interface (BCI) system is a potentially powerful new communication and control option for those with severe motor disabilities. BCI system translates brain activity into commands for a computer or other devices. Electroencephalography (EEG) is the most studied potential non-invasive interface mainly due to its fine temporal resolution, ease of use, portability and low set-up cost. Unfortunately, non-invasive implants produce a noisy signal because the skull dampens signals. Another substantial barrier to use EEG as a brain-computer interface is the extensive training required before users can work the technology.

Oddball paradigms are used by brain-computer interfaces to generate event-related potential (ERP) on targets selected by the user. The well-known P300 speller is based on this principle [1]. The main problem is to be able to detect ERP in a noisy electroencephalographic signal recorded on human scalp. Literature proposes a set of methods to detect ERPs based on averaging. Linear or nonlinear alignment is used to deal with the variable latencies in ERPs. A good alignment allows to obtain a better averaged ERP response, named "ERP template" here, and then to be able to detect ERP faster.

Section 2 of this paper presents a set of methods of template-based classifiers. Section 3 presents the results of these techniques on a well-known third BCI2000 competition data set acquired from a P300 speller. In section 4, we conduct the conclusion and give recommendation for future works.

## 2 Methods

Methods design for ERP detection use features extracted from the responses in the time domain. To extract features from the frequency domain after any transformation is difficult in this situation because ERPs are short-time events with a local peak. Here, we discuss two types of methods using one or several templates estimated explicitly and inexplicitly in the time domain respectively.

## 2.1 Classifiers based one ERP template

In oddball paradigm, one deviant rare stimulus is presented to subjects amongst a train of standard ones. Only the deviant stimulus induces an ERP. The most common evoked potential used in BCI is the P300. Its main characteristic is a positif latency that occurs around 300ms after the stimulus presentation. This leads us to derive a set of classifiers. All of them first explicitly estimate one ERP template by averaging the ERP responses in the time domain, and then use the distance between the response and the ERP template as the discriminant criterion.

**point-to-point averaging (P2P) classifier:** P2P classifier discriminates the ERP responses from non-ERP responses in a straightforward way: first it estimates one ERP template by simply averaging the measurements of ERP responses point-to-point, and then calculates the Euclidean distance between responses and the ERP template. For the P300 speller, the ERP response corresponds with the one producing the minimum distance within a set of column or row responses. Then the other responses are considered as non-ERP responses.

P2P classifier is simple for implementation. However the latency of the components of ERP varies with the external factors (such as target-to-target interval). Simply averaging the measurements point-to-point could blur the component. We then seek the solution of alignment before averaging. Such alignment includes linear alignment (such as, cross-correlation alignment) and nonlinear alignment (such as, dynamic time warping).

**cross-correlation averaging (CC) classifier:** Linear alignments can be done by shifting the measurements before the point-to-point averaging. One of the well-known linear alignments is cross-correlation. The best shift of one sequence in order to align it to the other ones is given by the one producing the highest correlation [2, 3]. CC classifier first searches the best shift for pairs of ERP responses before averaging them point-to-point to obtain one template. Then it calculates the cross-correlation between the testing response and the ERP template. The predicted ERP response corresponds with the one giving the maximum correlation value with the template. Such scheme accommodates variable linear latency. However, if one looks the ERP measurements more closely, it is found that the latency of the ERP component is nonlinear. In this case, the nonlinear alignment methods are better than the linear ones for template estimation.

**dynamic time warping averaging (DTW) classifier:** Dynamic Time warping distance is widely used in speech domain. It is able to deal with compression/expansion by comparing the distance of each point of the first sequence to every point of the second one using the information of the amplitude and/or the first derivative of amplitude of the signals [4]. Hence, DTW classifier first finds out the best nonlinear alignment path giving the closest distance, and then averages the ERP responses according to the alignment path to obtain one ERP template [5, 6, 7]. Finally, it recognizes the ERP responses as the one producing the minimum DTW distance.

## 2.2 Methods with several templates of ERP & non-ERP

On one hand, considering the complexity of ERP responses, one template could be insufficient to capture the variability of ERP components. On the other hand, one would like to introduce non-ERP templates into decision making process done by the classifiers. Clustering methods can produce several templates. The aim of clustering methods is to split a set of patterns into clusters as homogeneous as possible, i.e., patterns are gathered by similarity. A classical rule to assign a new pattern to a cluster is based on distances between the input pattern and templates. LVQ is the most famous algorithm in this category.

**learning vector quantization (LVQ):** The first version of the learning vector quantization algorithm proposed by Kohonen [8] is a supervised clustering method designed for classification. Input weights (IW) of connections linking inputs to each neuron is a vector $m$ with the same dimension than input $x$. Each neuron is assigned to a class according to the pre-defined layer weights (LW). A classic competitive step is used to determine which neuron is the closest one to the current input with the formula: $c = arg \min_i\{||x - m_i||\}$, and only the winner neuron is updated according to the following rules:

$m_c(t+1) = m_c(t) + \alpha(t)[x(t) - m_c(t)]$ if $x$ and $m_c$ belong to the same class

$m_c(t+1) = m_c(t) - \alpha(t)[x(t) - m_c(t)]$ if $x$ and $m_c$ belong to different classes

where $\alpha$ is learning rate. To apply this algorithm, one needs to decide the value of learning rate and the number of neurons assigned to each class. Further considering the multichannel recording of EEG measurements, if one simply merges features from each channel, the dimension of IW increases with the number of channels, and it thus causes memory problem in simulation. We then consider multichannel LVQ.

**multichannel learning vector quantization (mLVQ):** For multichannel LVQs, one LVQ model is created for each channel, and trained according to the above rules to update IWs. Thus, there are several ERP and non-ERP templates generated for each channels. In order to combine the templates in an optimal way, we further introduce an update scheme for LWs: $LW = (-||X - IW||)^{\dagger}T$, where $X$ is the input matrix and $T$ is the target matrix. The symbol $\dagger$ represents the pseudo-inverse operator. Such a solution of $LW$ is known as minimum norm least square solution. It is originally found being used in extreme learning machine (ELM) in [9].

# 3 Results

## 3.1 Wadsworth BCI data set

The results presented here are based on the P300 speller data set from the BCI competition III [10] [1]. The data set contains training and testing EEG data from two subjects. There are 85 letters for training and 100 letters for testing. For each letter, the recording consists of 15 epochs, and within each epoch, there are 12 flashings. For each epoch, a random permutation is chosen to highlight rows and columns. There is a 6x6 grid in this application containing 26 letters, 9 digits and one dash character. We are interested in the measurement window starting from the onset of the flashing and till one second, which consists of 240 samples at a sampling rate of 240Hz. In case the row/column is highlighted exactly where the target letter is located, we got the ERP response; otherwise the non-ERP response.

## 3.2 Experimental Results

### 3.2.1 ERP Templates

Figure 1 shows the comparison between the ERP templates estimated by P2P averaging, CC and DTW alignment averaging, respectively. Templates are built using the first 128 trials of the training set of subject A. Indeed, the algorithms using CC and DTW prefer to pair trials in a binary tree structure. The latency of the P300 component in the estimated templates by CC and DTW alignment averaging methods is different from P2P averaging. This is the effect of aligning the component in the time domain before averaging. And we found that the template estimated by DTW was smoother than by cross-correlation.

### 3.2.2 Classifier Comparison

Simulation is carried out to compare the performance of the template-based classifiers using P2P, CC, DTW and LVQ. The result of linear discriminant analysis (LDA) is also provided for its

[1] The data set is available at: http://ida.first.fraunhofer.de/projects/bci/competition_iii/

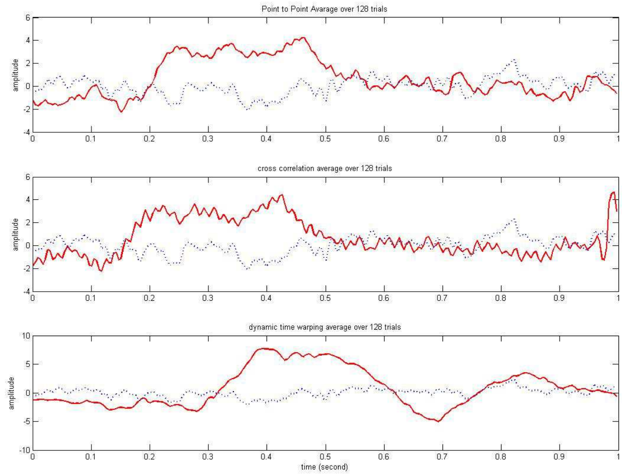


Figure 1: Comparison of ERP templates (red solid curves) and averaged background EEG activities (black dot curves) for responses recorded on channel Cz using training data from subject A. From top to bottom, templates are estimated by point-to-point averaging, cross-correlation and dynamic time warping alignment averaging, respectively.

simplicity and good performance in the application of BCIs. Table 1 shows the testing accuracy for subjects A and B. The results are based on the raw measurements from channel Cz (i.e., no pre-processing is involved). We used the full training set to estimate the ERP templates for P2P, CC and DTW; while for LVQ and LDA, they are used to estimate the parameters of the models. The full testing set is used to evaluate the performance of the classifiers. From the table, we can see that LVQ and LDA achieve similar performance and are better than the classifiers using P2P, CC, and DTW.

Table 2 shows the testing accuracy for subjects A and B based on measurements of all channels. Due to limit of memory and the large size of the input data, one needs to re-sample the signals. Thus, for Table 2, we first smooth the signal using a moving averaging filter with window size of 13, and then re-sample it with a down-sampling factor of 13. Further, multichannel version of LVQ is used here. We find that the performance of mLVQ and LDA are similar again and both increase as more channels are used for classification; while the performance of the explicit template-based classifiers using P2P, CC and DTW is bad.

From the above simulation results, we found that the performance of classifiers (such as LVQ), which take account of both ERP and non-ERP templates is better than those (such as P2P, CC, and DTW) which use only one ERP template. The optimization of the combination of the templates from different channels through assigning different weights is even more critical for improving the accuracy.

# 4 Conclusion and further work

BCI applications based on P300 ERP detection, such as the P3speller, need statistical or machine learning techniques with temporal features. ¿From a recording point of view, ERPs are graphic-elements. They have a specific waveform clearly different from EEG background activity. Thus, efficient modeling of the waveform can help to detect ERPs. Averaging techniques and clustering methods allow extracting one or several ERP templates. Averaging techniques using point-to-point averaging, cross-correlation and dynamic time warping are not efficient. They produce only one ERP template and suffer from two difficulties: first, responses are too noisy to easily distinguish ERP from non-ERP responses; second, according to the previous remark, they do not

| | | Method | | | | |
|---|---|---|---|---|---|---|
| Subject | Epoch | P2P | CC | DTW | LVQ | LDA |
| | 05 | 7% | 8% | 15% | 17% | 18% |
| A | 10 | 5% | 26% | 28% | 28% | 30% |
| | 15 | 3% | 21% | 39% | 43% | 41% |
| | 05 | 2% | 6% | 6% | 8% | 15% |
| B | 10 | 5% | 9% | 10% | 13% | 19% |
| | 15 | 3% | 16% | 15% | 21% | 26% |

Table 1: Experimental results based on channel Cz for subjects A and B. The percentage of correct classified letters contained in the testing set is shown regarding using the first 5 epochs, 10 epochs and all epochs.

| | | Method | | | | |
|---|---|---|---|---|---|---|
| Subject | Epoch | P2P | CC | DTW | mLVQ | LDA |
| | 05 | 9% | 7% | 7% | 47% | 45% |
| A | 10 | 13% | 6% | 12% | 77% | 78% |
| | 15 | 14% | 7% | 15% | 87% | 88% |
| | 05 | 3% | 2% | 4% | 72% | 76% |
| B | 10 | 3% | 3% | 3% | 91% | 92% |
| | 15 | 4% | 4% | 3% | 96% | 96% |

Table 2: Experimental results based on all channels for subjects A and B. The percentage of correct classified letters contained in the testing set is shown regarding using the first 5 epochs, 10 epochs and all epochs.

take into account the specificities of non-ERP responses to catch small differences between noisy ERP and non-ERP responses. The first difficulty exists for any classification method. But, our method mLVQ obtains similar results as LDA which is a reference technique for the P3Speller. It is because mLVQ takes into account the second remark. Thus, template-based classifiers can also obtain good results. Moreover, this approach can extract knowledge about the ERP waveform of patients.

In this article, we also evaluated the promising dynamic time warping distance. This kind of similarity measure is widely used in speech domain and align efficiently two sequences even if they differ each other from compression and/or dilatation on some segments. However, from our opinion, DTW suffers from the low SNR. The real EEG signals used in our simulation are too noisy. Many artificial peaks appear in the recording and thus it is difficult to obtain a strong alignment. In order to mitigate this difficulty, one may reduce the window of investigation to [250ms, 650ms] and constrain the window size of permitted distortion to short time. It should help to focus the attention on the critical zone.

In practice, it is also very interesting to reduce the number of channels in order to equip the patient with a compact system with only several necessary electrodes. Thus, channel selection will be investigated in a further work.

# References


[1] L.A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510–523, 1988.

[2] C. D. Woody. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Med. Biol. Eng.*, 5:539–553, 1967.

[3] D.G. Wastell. Statistical detection of individual evoked responses: an evaluation of woody's adaptative filter. *Electroencephalography and Clinical Neurophysiology*, 42:835–839, 1977.

[4] E. Keogh and M. Pazzani. Derivative dynamic time warping, 2001.

[5] T. Picton, M. Hunt, R. Mowrey, R. Rodriguez, and J. Maru. Evaluation of brain-stem auditory evoked potentials using dynamic time warping. *Electroencephalography and Clinical Neurophysiology*, 71(3):212–225, 1988.

[6] K. Wang, H. Begleiter, and B. Porjesz. Warp-averaging event-related potentials. *Clinical Neurophysiology*, 112(10):1917–1924, 2001.

[7] S. Casarotto, A. M. Bianchi, S. Cerutti, and G. A. Chiarenza. Dynamic time warping in the analysis of event-related potentials. *IEEE Engineering in Medicine and Biology Magazine*, 24(1):68–77, 2005.

[8] T. Kohonen. The self-organizing map. In *Proceedings of the IEEE*, volume 78, pages 1464–1480, 1990.

[9] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme learning machine: A new learning scheme of feedforward neural networks. In *Proceedings of International Joint Conference on Neural Networks*, volume 2, pages 985–990, Budapest, Hungary, 2004.

[10] D. J. Krusienski and G. Schalk. *Wadsworth BCI Dataset (P300 Evoked Potentials)*. BCI Competition III Challenge, 2004.